

ENCODE continues process to translate book of life

HudsonAlpha takes part in publishing genomic insights

September 4, 2012

EMBARGOED until noon CST, September 5, 2012

HUNTSVILLE, Ala. -- A milestone in genomic research achieved and shared by an international team working on the ENCODE Project, including scientists from the HudsonAlpha Institute for Biotechnology, means the task of interpreting an individual's genome and better understanding human disease risk is getting closer to realization.

The Human Genome Project—the sequencing of the first human genome that was completed in 2003—revealed about 3 billion base pairs comprise human DNA. In a way, the reading of those base pairs could be likened to learning the alphabet. Since 2007, the ENCODE Project has been taking the long strings of letters and discovering meaning from them by first identifying words, spaces and punctuation, then building sentences, constructing paragraphs and composing chapters. “The goal, in a sense, is to interpret and document our genetic makeup,” said Rick Myers, Ph.D., president and director of the HudsonAlpha Institute for Biotechnology.

The team of 442 researchers from 32 labs in the U.S., U.K., Spain, Japan and Singapore mapped more than 4 million regulatory regions where proteins interact with the DNA with specificity. This mapping significantly advances understanding of the precise and complex controls exerted on cell function.

Among important discoveries released today, researchers found that more than 80 percent of the human genome sequence is linked to biological function through regulation of our genes. “There are millions of additional elements encoded in our DNA, involving more letters of our genome than do the protein-coding genes,” said Myers. These elements help determine which sets of proteins, as well as the quantities of each of them, are made in each of cell.

In the last 10 years, researchers learned that 1 percent of the 3 billion paired letters in the human genome code for proteins. The data from ENCODE has further illuminated functional elements in the genome while also giving a better understanding of how these elements are orchestrated in complex networks. Because of the ongoing work of the ENCODE Project, scientists have new ways to anticipate and target future research.

“The research reported by the ENCODE Project today is a crucial step toward deciphering the language of regulation that underlies human biology,” said Myers.

Scientists at HudsonAlpha, in close collaboration with a group at the California Institute of Technology led by Barbara Wold, Ph.D., were major contributors to the ENCODE Consortium. One of their efforts was to look at the complex process of gene regulation and chromatin using a process called ChIP-seq, or chromatin immunoprecipitation sequencing. This process uncovers places in the human genome where a specific protein is bound, in a specific cell type. ENCODE scientists focused on proteins called transcription factors, known to bind near genes and regulate gene expression. “By cataloging all of the places where 119 of these factors bind our genome, we could see complex relationships between factors in different cells,” said Myers.

-more-

Most genes, Myers continued, do not have just on or off switches; rather, there is a detailed hierarchy regulating timing and functionality. Furthermore, genetic differences between people can affect the switches themselves. “Together, these massive datasets help us to better understand when and how disease occurs.”

That kind of understanding, Myers said, enables the creation of new ways to anticipate, diagnose, treat and prevent disease. “These proteins and the switches that control them are critical for fundamental, everyday biological activities -- metabolism, locomotion, reproduction, cognition, even aging. These components work together smoothly most of the time, but things can go wrong, leading to disease, poor responses to our environment and other problems.”

The ENCODE Consortium placed the resulting data sets as soon as they were verified for accuracy, prior to publication, in several databases that can be freely accessed by anyone on the Internet. These data sets can be accessed through the ENCODE Project portal (www.encodeproject.org)

The ENCODE maps allow researchers to inspect the chromosomes, genes, functional elements and individual nucleotides in the human genome. Myers added, “As a result of the ENCODE Project, thousands of researchers around the world -- in universities, medical schools, biotech companies, pharmaceutical companies, and research institutions -- now have this new interpretation of our genome, freely available from the Internet, and are using it to understand a wide range of basic and biomedical problems.”

The coordinated publication set includes one main integrative paper and five other papers in the journal *Nature*; 18 papers in *Genome Research*; and six papers in *Genome Biology*. To help anyone use and understand their data, the researchers also developed a novel way to follow any one biological theme through all of the papers; this can be viewed at <http://www.nature.com/encode>.

###

Editors:

Scientific inquiries – cgunter@hudsonalpha.org

Media inquiries – hralston@hudsonalpha.org

The HudsonAlpha Institute for Biotechnology in Huntsville, Ala, is the cornerstone of the Cummings Research Park Biotechnology Campus. The campus hosts a synergistic cluster of life sciences talent - science, education and business professionals - that promises collaborative innovation to turn knowledge and ideas into commercial products and services for improving human health and strengthening Alabama’s progressively diverse economy. The non-profit institute is housed in a state-of-the-art, 270,000 square-ft. facility strategically located in the nation’s second largest research park. HudsonAlpha has a three-fold mission of genomic research, economic development and educational outreach.